Defense Threat Reduction Agency
8725 John J. Kingman Road, MS
6201 Fort Belvoir, VA 22060-6201

DTRA-TR-13-54

TECHNICAL REPORT

Pre-incident Analysis using Multigraphs and Faceted Ontologies

Approved for public release; distribution is unlimited.

August 2013

HDTRA1-03-D-0009

T.P Caudell, et al.

Prepared by:

OVPR / University Strategic Partnership
MSC02 1660, 1 University of New Mexico
Albuquerque, New Mexico 87131

DESTRUCTION NOTICE:

Destroy this report when it is no longer needed.
Do not return to sender.

PLEASE NOTIFY THE DEFENSE THREAT REDUCTION AGENCY, ATTN: DTRIAC/ J9STT, 8725 JOHN J. KINGMAN ROAD, MS-6201, FT BELVOIR, VA 22060-6201, IF YOUR ADDRESS IS INCORRECT, IF YOU WISH THAT IT BE DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
00-08-2013	Technical	07/21/210 - 04/30/2012

4. TITLE AND SUBTITLE

Pre-incident Analysis using Multigraphs and Faceted Ontologies

5a. CONTRACT NUMBER
DTRA01-03-D-0009-0026

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

T.P Caudell (PI), G. Luger (CoPI), F. Gilfeather, M.J. Healy, S. Smith, D. Weinberg, J. A. Turner, M. D. Turner

5d. PROJECT NUMBER
2

5e. TASK NUMBER
26

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

OVPR / University Strategic Partnership
MSC02 1660, 1 University of New Mexico
Albuquerque, New Mexico 87131

8. PERFORMING ORGANIZATION REPORT NUMBER

OVPRED 798B

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Threat Reduction Agency/RD-BAT
8725 John J. Kingman Rd.
Ft. Belvoir, VA 22060-6201
PM/ James Reed

10. SPONSOR/MONITOR'S ACRONYM(S)

DTRA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

DTRA-TR-13-54

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Situational awareness requires acquisition of meaningful and reliable information. In any number of operating environments, large streams of raw information must be analyzed and processed by agencies that range from law enforcement to emergency services during a crisis. This research focuses on information related to strategic intelligence collection and analysis. Reports obtained by such processes reveal only pieces of the situational picture – it is the combination of many reports (from different analysts and sources) that potentially reveal the underlying picture. Decision makers will benefit greatly from methods that organize information into new semantic perspectives different from that in which it was collected. This research investigates the organization of context specific information into semantic graphs and the merging of the semantic graphs into a multigraph to create a faceted ontology. This organizes the viewpoint-specific semantic graph structures into a more readily interpretable, robust, perspective neutral representation. The simpler semantic structures are collected from various sources focusing on, for example, socio-cultural networks, geo-spatial distributions, or threat scenario trees.

15. SUBJECT TERMS

algorithmics multi-graphs structure onltologies

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					James Reed
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	40	**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			703-767-8793

CONVERSION TABLE

Conversion Factors for U.S. Customary to metric (SI) units of measurement.

MULTIPLY → BY → TO GET

TO GET ← BY ← DIVIDE

angstrom	1.000 000 x E -10	meters (m)
atmosphere (normal)	1.013 25 x E +2	kilo pascal (kPa)
bar	1.000 000 x E +2	kilo pascal (kPa)
barn	1.000 000 x E -28	meter^2 (m^2)
British thermal unit (thermochemical)	1.054 350 x E +3	joule (J)
calorie (thermochemical)	4.184 000	joule (J)
cal (thermochemical/cm^2)	4.184 000 x E -2	mega joule/m^2 (MJ/m^2)
curie	3.700 000 x E +1	*giga bacquerel (GBq)
degree (angle)	1.745 329 x E -2	radian (rad)
degree Fahrenheit	$t_k = (t^\circ f + 459.67)/1.8$	degree kelvin (K)
electron volt	1.602 19 x E -19	joule (J)
erg	1.000 000 x E -7	joule (J)
erg/second	1.000 000 x E -7	watt (W)
foot	3.048 000 x E -1	meter (m)
foot-pound-force	1.355 818	joule (J)
gallon (U.S. liquid)	3.785 412 x E -3	meter^3 (m^3)
inch	2.540 000 x E -2	meter (m)
jerk	1.000 000 x E +9	joule (J)
joule/kilogram (J/kg) radiation dose absorbed	1.000 000	Gray (Gy)
kilotons	4.183	terajoules
kip (1000 lbf)	4.448 222 x E +3	newton (N)
kip/inch^2 (ksi)	6.894 757 x E +3	kilo pascal (kPa)
ktap	1.000 000 x E +2	newton-second/m^2 (N-s/m^2)
micron	1.000 000 x E -6	meter (m)
mil	2.540 000 x E -5	meter (m)
mile (international)	1.609 344 x E +3	meter (m)
ounce	2.834 952 x E -2	kilogram (kg)
pound-force (lbs avoirdupois)	4.448 222	newton (N)
pound-force inch	1.129 848 x E -1	newton-meter (N-m)
pound-force/inch	1.751 268 x E +2	newton/meter (N/m)
pound-force/foot^2	4.788 026 x E -2	kilo pascal (kPa)
pound-force/inch^2 (psi)	6.894 757	kilo pascal (kPa)
pound-mass (lbm avoirdupois)	4.535 924 x E -1	kilogram (kg)
pound-mass-foot^2 (moment of inertia)	4.214 011 x E -2	kilogram-meter^2 (kg-m^2)
pound-mass/foot^3	1.601 846 x E +1	kilogram-meter^3 (kg/m^3)
rad (radiation dose absorbed)	1.000 000 x E -2	**Gray (Gy)
roentgen	2.579 760 x E -4	coulomb/kilogram (C/kg)
shake	1.000 000 x E -8	second (s)
slug	1.459 390 x E +1	kilogram (kg)
torr (mm Hg, 0^0 C)	1.333 22 x E -1	kilo pascal (kPa)

*The bacquerel (Bq) is the SI unit of radioactivity; 1 Bq = 1 event/s.

**The Gray (GY) is the SI unit of absorbed radiation.

-

T.P Caudell[1] (PI), G. Luger (CoPI)[1], F. Gilfeather[1], M.J. Healy[1],
D. Weinberg[2], S. Smith[3], J. A. Turner[1,4], M. D. Turner[4,5]

[1] University of New Mexico
[2] Practical Risk, LLC
[3] Los Alamos Visual Analytics
[4] New Mexico Mind research Network
[5] Conjectural Systems

Table of Contents

1.0 Motivation

1.1 Objective.

The object of this project is to advance the theory and algorithmics of new knowledge structures that are simultaneously more coherent in modeling real-world knowledge and more accessible to data analysts from many points of view, providing access to details while maintaining appropriate context. Our approach involves advancing and integrating theories of multi-faceted semantic structures based in category theory, theories of faceted ontologies using multi-graphs, theories of case-based inference within these structures, and theories of metaphorical understanding and representation.

1.2 General Background

In the context of human systems that operate toward a common goal, many attributes of both humans and the geo-temporal environment they operate in create connections, dependencies, and interdependencies. Coupling these to state variables that are essential to the human action, but interact with it, the number of connections (links) between relevant intersects (nodes) becomes almost intractable. The Internet provides a simple example Barbasi(2003). In the realms of situation awareness and decision analysis, far more focused capabilities are needed to sort through many connections of interest and yield results that provide visibility and support the intuition of the intelligence analyst. Such is the challenge faced by intelligence services, law enforcement (McCue, 2007), and emergency managers.

One arena where such issues can be clearly illustrated is the problem faced by intelligence agencies concerned with WMD/E development and attacks. Information is ingested by intelligence analysts in a more-or-less chaotic fashion. The information is incomplete, only occasionally verifiable, and rarely actionable (Grabo, 2004; Clark 2007). Typically, it is highly fractionated, of unknown veracity, and temporally displaced. With all the unknowns and the highly variable nature of attributes of the information, conventional graphs are ill suited to represent or portray a useful picture to an analyst concerned with the state of WMD/E development globally (or perhaps locally).

Within the Intelligence community, information management and its manipulation using advanced software is one of the primary tasks of analysts (Khalsa, 2004). Most analysts are specialists within a specific area, and are generally organized by country, specific technical fields such as nuclear, chemical, and biological, or topical such as terrorist groups or individuals. Everyday new information across a multitude of topics is provided to analysts. The information comes in a range of classification levels from open-source human terrain data to highly classified and compartmentalized, and in many types: 1) signals intelligence (SIGINT), 2) imagery intelligence (IMINT), 3) measurement and signature intelligence (MASINT), 4) human-source intelligence (HUMINT), 5) open-source intelligence (OSINT) and 6) geospatial intelligence[1]. The analyst must access, parse, and correlate each of these types of information on a daily basis to provide an analysis of the information, including the accumulated evidence for a case. In addition,

[1] http://www.intelligence.gov/2-business_cycle2.shtml.

the analyst often provides feedback to the collecting agencies on the type of information that is needed to help fill some of the gaps. An analyst may use sophisticated tools to find patterns within the information or may use rudimentary computer software such as word processing and spreadsheets to sort the information. A standard approach has not been established within the intelligence community with the exception of ensuring the information is tagged back to the original source.

Due to the large volumes of information on any one topic, an all-source analyst will typically focus on a specific topic or subtopic. Analysts first complete an extensive review of all past reporting and analysis on the topic in order to establish a baseline that is updated daily with new information. The analyst uses all of the six types of intelligence as needed and available for analysis and correlation. This can include imagery, video, maps charts, and other forms of information. For example, an analyst reviewing trends in nuclear diversions and their application to current nuclear threats creates a series of cases for collecting and correlating information on a case-by-case basis. This is a simple way to organize the information and establish confidence regarding the level of information and overall trends within nuclear diversions. This may include pictures of individuals, nuclear materials that have been apprehended, videotaped interviews, court files, and charts of the materials' compositions and trace elements. The individuals and groups involved constitute a second level of information to correlate for evidence of established patterns over time indicating that a specific terrorist organization, organized crime group, or country has been systematically seeking to obtain nuclear materials. This often requires an interface between multiple individuals and/or organizations to piece together larger patterns between different types of data, and some sort of framework within which the evidence can be accumulated. Automating this process requires the ability to first identify reports of interest based upon a natural-language processor that helps identify key words and phrases. This information can then be sorted in light of existing information organized by specific scenarios, threats, groups, persons, or cases to determine whether it adds knowledge to the existing information. An analyst must review the information to ascertain if the data is relevant and to provide an assessment of the confidence level that should be assigned to the source, the data, and its applicability to a specific location within an overall chain of activities or scenarios. This is a difficult task and requires analysts with an extensive background to assess the reporting. No automated tools currently exist to help analysts sort and connect critical data beyond what they can do manually and mentally.

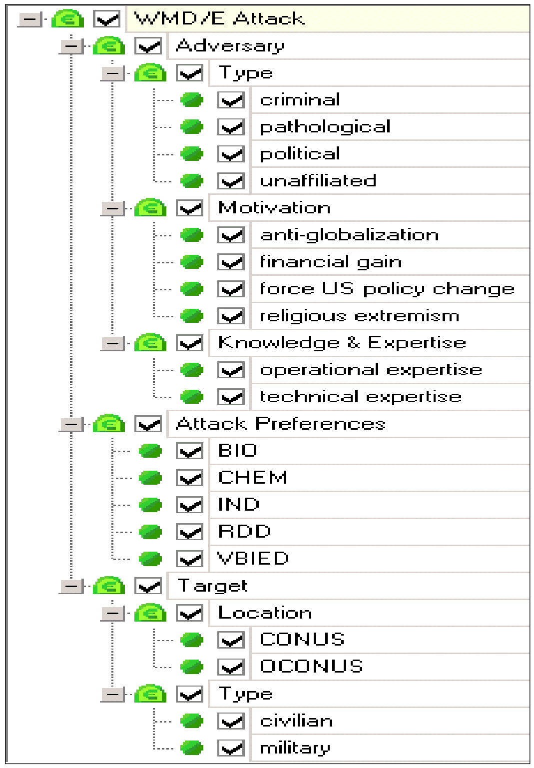

Figure 1. In simple tree form, the components of a WMD/E attack. Essentially, they consist of three major parts (or facets): the adversary, the adversary's chosen type of attack, and the type and location of the target.

Expressed in simple textual form, a paradigm for someone using a WMD/E has the following parts:

> "A motivated adversary with sufficient skill (or a method to obtain the skills) and resources (or a way, legally or illegally, to acquire those resources) must acquire a sufficient quantity and quality of whatever constitutes the critical component (without which the weapon would not exist), weaponize the component, acquire and test all weapon components, assemble and transport the device, and release it."

Figure 1 shows the major components of this series of activities leading to possible WMD/E attack. The format is that of a tree structure extracted using a software tool provided by Logic Evolved Technologies (Eisenhawer, 2009). Each component of the tree can be broken into many more sub-components, then further into sub-sub-components. The activities would be noticed by intelligence collection methods; for well-specified scenarios, these can be represented as a conventional tree or graph. Unfortunately, such specificity is almost uniformly lacking in creating true situational awareness with sufficient fidelity to provide actionable information.

In the WMD/E example, adversaries constitute an immense collection of individuals represented within one or more graphs. The adversary's beliefs, education, skills, social network, and goals all play into the type of attack he/she might attempt. Further, the list

of skills required to perform such an attack is not monolithic; it forms a broad spectrum ranging from finance to surveillance. Each step in this paradigm can be cast as a near-endless series of links and nodes in a graph. Importantly, there can be specific nodes and/or links not unique to a single attack paradigm, instead being shared among many graphs.

In addition, information or evidence coming in may in no way indicate to which graph it belongs. For example, explosives (along with all the skills needed to formulate, transport, and use them) are the fundamental critical ingredients of a vehicle-borne improvised explosive device (VBIED) such as the one used by Timothy McVeigh in Oklahoma City. However, explosives are also critical in an implosion-based improvised nuclear device (IND), or can be used as a dispersal mechanism for a radiological dispersal device (RDD) attack. Hence, the common convention of dealing with information used in situational analysis is far beyond the ability of traditional trees or graphs. Something more powerful is needed to effectively coalesce the various facets of this highly diffuse cloud of interconnected bits into a useful knowledge structure for the analyst. We propose that faceted ontologies will serve this need.

2. Theoretical Frameworks

2.1 Background on the theoretical approach-ontologies and category theory

The main thread of our theoretical approach is based upon the expression of faceted ontologies within the mathematical discipline of category theory, and with collaboration from AI software technologies in areas such as case-based reasoning. Category theory provides a mathematically rigorous framework for conceptual modeling and collaboration among knowledge technologies. As the name indicates, faceted ontologies provide a view of knowledge for from multiple viewpoints, a way of cross-indexing information with semantic depth. We shall investigate the potential of this idea as a comprehensive framework to support intelligence analysts.

Ontology can be described as a categorization of that which is perceived to exist—a way of answering the question “What is there?”. For example, Figure 1 attempts to describe the “what is there” in a WMD/E attack. By a category, Aristotle, who first organized ontology as a field of study, meant a class or collection of things all of one type, where typicality is determined by a mental representation or concept. Seen as a system of interrelated categories and their associated concepts, an ontology is a statement of being—a view of how things exist by virtue of their relation to others—and the concepts of typicality express this in the form of characteristic descriptors. A faceted ontology is a knowledge structure in which a thing may be represented in multiple categories, and is similar to a poly-hierarchy. For example, if there were an organization ontology representing *who might do it*, and an attack type ontology representing *what they might do*, the faceted ontology resulting from the merger of these two would simultaneously represent knowledge about *who might do what*.

Community ontology development in analysis arenas such as those described in the previous section is a rather recent activity, and while there is consensus on the issues involved few effective methodologies have been proposed. Unlike ontology applications in business that support fairly well specified activities, applications in intelligence analysis must support generative activities that by nature are not fully specifiable. Ontologies must support an open ended analytical process, not a discrete set of pre-defined activities. Hence, in analytical ontology development a major task is simply understanding the problem space and defining an abstract conceptual model that will transform “knowledge in the wild” into computable knowledge representations.

A faceted ontology integrates declarative information (statements about information content) across multiple perspectives in a format that enables the automatic exchange of data, analyses or other technical resources. Prior to producing a faceted ontology there must be community agreement on how those multiple perspectives can be integrated. Any approach that articulates and makes explicit an aspect of the knowledge domain such that others can understand it enables the process of negotiating linkages between perspectives (Lee, 2007). In the existing approaches, this is enabled through production of other kinds of knowledge artifacts. Any formal mathematical and computational constraints that would be required to update and maintain a coherent ontology are loosened in these informal, predecessor artifacts. Informal approaches include term lists, user-generated taxonomies, and concept maps

The transition between informal material artifacts and formal ontologies is perhaps the most problematic aspect of ontology building. Ontologies are a language for representing knowledge, and like any language there are syntax rules that must be followed. Additionally, because tools that operate on ontologies depend on logical consistency, both syntax and logic must be rigorous. Currently, the transition between informal community knowledge building and formal ontology building remains a complex human activity, though in some cases the transition is fairly simple. For instance, categorization of terms in a taxonomy or concept map can sometimes translate to an ontology in a straightforward way. Unfortunately, categorization of terms is necessary but insufficient for developing the kinds of linkages between concepts that we are targeting in this proposal. Understanding translation mechanisms between informal and formal knowledge representations through intermediate artifacts is a key research area in this field. Another problem is the ambiguities often present in incoming information; this ambiguity is only exacerbated by it being summarized by words and phrases in natural language. The latter is fraught with its own inherent ambiguities in representing knowledge outside the semantic realm (the situational context) from which the knowledge was derived. Hence, something deeper than labeling with terms is needed. In particular, the trend in concept representation is toward those with deeper explanatory content, notably logical theories. (Medin, 1989).

In much modern work (Tversky, 1977; Malt & Johnson, 1992; Murphy & Wisniewski, 1989), a concept is often represented as a set of "features" such as shape, color, or function. Often ignored is the fact that the relationships between features are as important a determinate of a concept as are the features themselves. Because the items in a category are of a type as expressed in the accompanying concept, relationships between category members are also important. Thus, not only are relationships between features important in determining category membership, but relationships between category members exists by virtue of the conceptual descriptors of a category. The relationships reflect the fundamental knowledge expressed in the concept that explains why something is a member of the category. It is the notion of structure associated with categories that suggests the use of category theory in conceptual analyses, for it is the mathematical theory of structure.

The relationships within many categories form a hierarchical structure. Often, this takes the form of a hierarchy of abstractions. Categories themselves are related by superordination; 'furniture' is a superordinate for 'chair', 'table', etc. Because its concept imposes fewer constraints on membership (furniture can be sat upon or used to support other items whereas a chair or table has a more specific function along with a physical description), the superordinate category has relatively more members. Notice also that the superordination relation is transitive, which makes superordination among categories is a special case of a compositional relation. Hence, the structures of interest in the proposed effort are compositional systems of relationships. Superordination can also be seen as a mapping between structures. Because each member of a category can be associated with a unique member of a superordinate category, the superordination relation on a pair of taxonomic categories can be seen as an example of an important type of mapping called a function. But the mapping is more than a function. More generally, a mapping between categories is an association of both items and their relationships and preserves the compositions. This makes it a structure-preserving mapping.

A faceted ontology separates its categorization into multiple categorizations of the same items from different viewpoints. For example, one taxonomic categorization of animals can be based upon physical properties such as morphology, another upon habitats, another upon behaviors, and so forth. To be useful, the items in facets must be related so that the same item can be seen from the different viewpoints. But it is important also that the relational structures of the facets be related in such a way that either there are links between the expressions of two related items in the different facets or else the missing links are identified and somehow compensated for in the system for operating upon and using the ontology. Also, as with any ontology, a faceted ontology has data-, process-, or situation-specific contexts to which it applies. For example, incoming data for an evolving situation must be analyzed in the context of existing knowledge. This results in a structuring of the data according to an existing ontology together with a synthesis of new knowledge by combining ontology concepts with the data; the new knowledge then expands the ontology. Here again, the notion of structure is fundamental.

Category theory is a recent branch of mathematics based upon the view that structure is important in categories and in relationships between them. Some comprehensive references are (Adamek et al., 1990; Crole, 1993; Lawvere & Schanuel, 1995; Mac Lane, 1971; Pierce, 1991). Relationships between categories are the key notion, expressed in terms of structure-preserving mappings. A category consists of entities of some kind, called objects, and relationships between them, called morphisms, together with a law of composition for the morphisms: the composition of $f: a \rightarrow b$ in a category C with $g: b \rightarrow c$ (also in C; notice that b is the "head" of one arrow and the "tail" of the next) is a morphism $g \circ f: a \rightarrow c$ in C. The composition operation, $\circ$, satisfies two important laws. However, the important thing to know for this discussion is that in almost all categories, it often happens that two compositions involving different morphisms but with the same beginning and end objects can be the same. For example, in addition to the composition just illustrated, it can happen that there is $h: a \rightarrow d$ and $k: d \rightarrow c$ also and that $(k \circ h: a \rightarrow c) = (g \circ f: a \rightarrow c)$. This fact is of fundamental importance, for it defines the notion of a commutative diagram, which is like a graph extracted from a category but with compositions of links and with some compositions being one and the same link. This yields a structural law and is fundamental in mathematical semantics. It is worth noticing at this point that a category has an underlying directed multigraph structure. This is fortunate, for the fact that multigraphs have no notion of composition can be remedied by reformulating them as categories.

Functors are structure-preserving mappings between categories in that a functor preserves composition. The importance of this is that a functor maps commutative diagrams in one category to commutative diagrams in the other. This is a transportation of semantic information between categories. Finally, there are many levels of structure in this formalism. For example, functors have a composition, and they are in fact morphisms in categories of categories. Natural transformations relate the transport of semantic structure in two functors, so serve as a mapping between functors. Again, there is a composition operation, leading to the important notion of functor categories.

There is, in fact, a wealth of mathematical machinery here. In the area of categorical logic (Lawvere, 1963; J. A. Goguen & Burstall, 1984; Meseguer, 1989; J. A. Goguen &

Burstall, 1992; Crole, 1993), this machinery has been proven in a variety of research areas with some resulting applications. Vickers discusses a formalism in a categorical logic and its accompanying model theory that has been applied in the area of formal specifications for software and for analyzing database semantics (Vickers, 1992). Johnson and Rosebrugh (2001) describe an approach for ontology formalization that has proven particularly effective in solving problems for enterprise information systems (Colomb, Dampney, & Johnson, 2001); for example, it has provided a new and more general theory-based but practical solution method for the “view updating” problem in database management (Johnson & Rosebrugh, 2001). Related uses of category theory are in system theory (J. Goguen, 1973) and again categorical logic in software synthesis (Burstall & Goguen, 1980; Jullig & Srinivas, 1993; Williamson & Healy, 2000), the mathematical study of biological systems (Rosen, 1958; Baianu, 1987; Ehresmann & Vanbre- meersch, 1997; Gust & Kuhnberger, 2005; Healy & Caudell, 2006a), and the formalization of ontologies (Uschold, Healy, Williamson, Clark, & Woods, 1998; Dampney, Johnson, & Rosebrugh, 2001). Categorical logic provides a vehicle for the formalization of ontologies with mathematical rigor. Entities of all types can be represented by variables and constants in closed symbolic formulas called sentences that express information about them. Sentences about the same entities can be grouped into theories, a way of formalizing concepts. Theories in categorical logic are accompanied by a model-theoretic foundation that allows an analysis of the instances, such as situations involving the entities that satisfy the sentences of a theory. The sentences of a theory state constraints upon its instances, called models. The models, in turn, form a category based upon relationships among the models. This provides a structure on and within the models—classes of entities, functions mapping between classes, and sub-classes defined by predicates. Ontologies can be formalized as theories, but a more powerful formulation is as categories of concepts (theories), with contexts represented by their model categories.

One approach to working with faceted ontologies is through semantic alignments between ontologies. A semantic alignment between ontologies expresses associations between their terms. This can be expressed in morphisms and categorical constructs based on morphisms. An example of this is the information flow (IF) methodology of Schorlemmer & Kalfoglou (2005), which derives ultimately from Goguen and Burstalls' institution theory (1992). In Zimmermann et al. (2006), alignments are studied via limits and colimits in categories of spans, which are more abstract in that they do not specify a particular type of mathematical structure such as formal logic for expressing ontologies. This has the advantage of allowing different ontology conceptualizations and contexts to be combined and supports collaboration.

2.2 Background on AI technology

Many complex application domains require the integration of multiple representational schemes, where each representational scheme captures different aspects of the situation being assessed. For example, a speech analysis tool can map the components of telephone communications into appropriate slots within a larger scenario. The slots are features specific to different aspects of scenarios, such as location, conversation subject matter, etc. The different representational schemes must be merged to address the full task of situational awareness. The schemes, whether rule systems,

constraint trees, traditional data base relations, web-scraping spiders, software for assessing graphic data, or simply large components of traditional computer code can each be represented by an ontology. Ontology merging requires a set of sophisticated software tools and support. The leaf node of a constraint tree in one ontology might contain the critical information for generating a new web query in another ontology and also for suggesting a search of a specific data base captured in yet a third ontology. Furthermore, the results of the data base query might need to be fed back, as a new constraint, into the original tree based constraint ontology. Merging software-encoded ontologies is a much more sophisticated task than simply combining the various combinations of code together into a larger system. The "glue" software for ontology intercommunication must "know" the structure and interfaces for each of the component ontologies

There are currently several software products available for building software systems for sophisticated ontology merging. OWL is one of the most pervasive tools (Mika, P., Oberle, D., Gangemi, A. and Sabou, M. 2004). All existing tools have strengths and weaknesses that must be resolved before they can serve the purpose of intelligence analysis. A major work item will be to investigate the unification of different representational schemes via categorical logic.

A second component that will be investigated for situation assessment is the use of case-based triggering and retrieval software. The key to this approach is that "cases" make up the primary top-level representation structure for situation assessment. The case data structure is a complex record of components that together describe a situation. In criminal law, a case includes the crimes, the date and time of each crime, the accused, a victim's list, witnesses, prior information on the suspect, the appointed judge, and related information such as video recordings, pointers to evidence, etc. in a particular criminal scenario. Similarly, in terror analysis, a case contains the relevant features of a situation. Importantly, a case often is only partly instantiated. For example, the evidence might not be fully catalogued, or a judge not yet appointed. Nonetheless, the case is the vehicle for describing this criminal situation, and in a software system it may be associated with procedures for relating it to other similar cases, or to other cases of the same suspect, judge, etc.. Additionally, when the case achieves a critical level of data or urgency, a software "trigger" can alert an observer. Existing case technology supports the representation of multiple scenarios in, for example, a database of collected cases. Some of these cases can be labeled as "cases of interest" while others provide background. Although the case-based reasoning tools have been explored primarily in the development of legal cases, the overlap with the development of terror scenarios is straight forward. Further information on case based technology may be found in (Luger, 2009).

In the field of information organization and extraction, an ontology is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. It is used to reason about the entities within that domain, and may be used to describe the domain. An ontology provides a shared vocabulary, which can be used to model a domain that is, the type of objects and/or concepts that exist, and their properties and relations. Information retrieval using ontology engineering techniques automatically extracts structured and categorized information that is contextually and semantically well defined from a specific domain from unstructured

machine-readable documents. These techniques are useful for richer extraction of information from sources that are a combination of structured and unstructured data. [Gangemi, 2005].

2.3 Visualization of Faceted Ontologies as Multigraphs

A key issue in this project is representing the knowledge structures is such a way that analysts can comprehend their meaning. This involves the visualization of these potentially complex structures encoded as large-scale multigraphs. The goal of an interactive visualization tool of this nature is to be able to tease out of a large, and apparently ambiguous structure in a large multi-relational graph, contingent structures based on particular preferences. The canonical example to consider would be the natural circumstance where intelligence is organized according to how it is gathered and conventionally understood. A natural way humans tend to organize information is in simple trees. It may be a subsumption hierarchy such as geographical regions (the city of X is within the province of Y in the State of Z in the Nation of Q on the Continent of P) or a taxonomy (the XYZ brotherhood is a subgroup of the ABC group who are HJK terrorists). Real world events tend to cross, inform, or connect many of these simple hierarchies into a more complex structure. The nodes in the graph can have many related or complementary properties, describing people, places, things, events, ideas, etc. The edges, as well, may have a wide range of properties, describing the relationships between the nodes such as, "is near", or "is an example of" or "has the property of" or "has acquired", etc. Visualization research has shown that certain representations of complex hierarchical information can enhance human comprehension in programming and data mining application domains. This will add considerable value in combination with the other theoretical and algorithmic approaches listed above in the area situational awareness.

3. General Approach

We propose to investigate the use of categorical model theory for the analysis and synthesis of faceted ontologies in the context of new and existing data. A faceted ontology separates its categorization into multiple categorizations of the same items from different viewpoints. But it is important also that the relational structures of the facets be related in such a way that either there are links between the expression of an item in different facets or else the missing links are identified and somehow compensated for in the system for operating upon and using the ontology. Also, a faceted ontology has data-, process-, or situation-specific scenarios to which it applies. For example, incoming data for an evolving situation must be analyzed in light of existing knowledge, which is expressed in the ontology. This can be regarded as a structuring of the data according to an existing ontology together with a synthesis of new knowledge by combining ontology concepts with the data, thereby expanding the ontology.

There exists a large category whose objects are all categories of interest and whose morphisms are functors. This category is augmented by the presence of natural transformations—morphisms between functors. This provides the machinery for a mathematically rigorous "interoperability" between categorical model theory and other formalisms, such as graphs (including semantic graphs), petri nets, tree structures and so forth, and technologies such as existing systems for multigraph visualization, ontologies and case-based reasoning. This machinery is central to this effort.

A major research challenge is the development of mathematical constructs general enough to model arbitrarily complex knowledge structures while being flexible enough to support multiple organizations of the knowledge for analysis. A secondary research challenge is how to populate these structures incrementally as data is gathered. By modeling these structures as faceted ontologies, whose facets reflect the natural way data is gathered or discovered in a context, we hypothesize that the resulting knowledge structure will most accurately reflect the underlying situations from which the data is derived. The structures appropriate for organizing the data as it is gathered does not necessarily reflect the structure of a data analyst's questions as they try to extract specific scenarios from the data. The rich structures of categories, functors and other categorical constructs provide a mathematical model for mapping between constructs that formalize multiple contexts, those in which data is collected and the scenarios analysts are evaluating.

Inference within the faceted ontologies will be studied with both classical AI methods and category-theoretic constructs. Case based representations of information will be recast in categorical terms to study mathematically the use of case-based reasoning methods with faceted ontologies. To test the formal theoretical understanding we anticipate coming out of this work, we will create new algorithms that instantiate the mathematical processes and analyze them with respect to their computational complexity and performance to achieve a characterization of scale-up to realistic problem scenarios. These empirical findings will inform the theoretical research to aid in the discovery of new mathematical methods to address any issues uncovered..

We also propose to extend several lines of our existing research in the visualization of complex graphs. The first is visualization of graphs in three or more dimensions. Most

graph layout research is limited to 2D. There are important features of 3D and higher layout that are desirable for this work (less likely to tangle, more compact by the 3/2 power, inherently multi-view by navigating around and projection of shadows, etc). The second is interactive graph layout intrinsically supports exploration of graph properties both by observing an evolving layout and by deliberately disturbing the layout and observing the results. Similarly, the generalized extension of force-directed layouts we are developing is suitable for interactive use to expose hidden features and relationships of graphs. We have also developed various techniques in semantic zooming that we intend to study as semantic filtering or lensing. Our novel techniques for viewing high-dimensional systems as projections in lower dimensions have not yet been studied in the area of multigraphs, and was in this project. All of these techniques may be applied separately or together. Again, category theory provides an overarching framework for collaborations that exploit the favorable aspects of graph visualization and other technologies.

The new knowledge developed by this research should provide decision makers with powerful methodologies to translate the semantic context from one decision maker to another, or from data gatherers to data analysts. Information gaps in one decision maker's domain can be completed by information in another domain linked through ontology merging. This research responds to the topic G request for conciliating and deconflicting data, and falls into Technology Readiness Levels 1 and 2

Summary.

Year #1 (Year 2010-11)

Task 1: Formal Mathematical Theory of Faceted Ontologies

Task 2: Case-based Inference in Faceted Ontologies

Task 3: Visualization of Faceted Ontologies

Detailed Tasks.

i. **Task 1: Formal Mathematical Theory of Faceted Ontologies.** a) Category theory will be used as a mathematical approach to formalizing ontology mergers and projections, b) An incremental application of category theory to first the merger problem, then projections, then inference.

ii. **Task 2: Case-based Inference in Faceted Ontologies.** a) Case based representations of information will be cast into discrete categories and categorical completion will be studies as a method to understand how the components of the cases can be assembled into full scenarios, b) An incremental study of case based reasoning and formal methods that apply to them.

iii. **Task 3: Visualization of Faceted Ontologies.** a) To extend our existing research in the visualization of complex graphs to multigraphs representing faceted ontologies, b) Search for existing algorithms, analysis, design and testing of new algorithms.

4. Results

4.1 Ontology Test Case Development

To serve as an example of concepts, relationships, and the complex types of knowledge for the faceted ontologies, we developed a foundation for an ontology to represent adversary groups and their intentions, classification of their weapons and attack types, and the ability to represent the relationship between the outcomes of an attack and the various recognized intentions of the adversary group. This Adversary-Intent-Target (AIT) model focuses on structuring knowledge to allow reasoning about which groups would be likely to choose what kinds of weapons to perform which kinds of attack. The AIT model is a generalizable and extensible system for organizing the relevant information, serving as a preliminary ontology within a larger computational system.

The full report resulted in a white paper available through the Department of Electrical and Computer Engineering Technical Reports, http://hdl.handle.net/1928/13714. A summary of the work is included here. The software tools being developed require a semantic ``grounding," that is, a controlled vocabulary of terms (words) with a fixed set of relations on the terms of that vocabulary that enforce a logical structure. Together these features form an ontology, in this case an ontology for terrorism research. With such an ontology in place, unthinking machinery can do a wide variety of (seemingly) intelligent reasoning tasks while still preventing the results from becoming semantic gibberish.

The current standard for representing ontology terms and their relationships on the semantic web is the Web Ontology Language or OWL. (See http://www.w3.org/TR/owl2-overview/ for details.) Specifically we use OWL2 for development of AIT. OWL is a family of languages with differing levels of logical expressiveness. We make use of OWL2 Full, in principle, but the bulk of our work is at the level of OWL2 DL (Description Logic), a restricted sublanguage of OWL2 Full with better computational properties.

The AIT model begins with a model statement regarding a terrorist attack (Fig. 3). This statement is a simple sentence in natural language:

A terrorist attack occurs when an adversary, with intent and capability, uses a weapon against a target.

This statement expresses a particular point of view (POV) toward terrorist attacks, and any POV implies a corresponding ontology. What we do in the AIT modeling process is develop the appropriate ontology for breaking up the world-at-large into parts and relations that reflect the assumptions embedded in this model statement. Part of the larger project for the associated computing research this AIT model was developed to support, is to allow multiple, overlapping ontologies to coexist and exchange data despite their differences in POV.

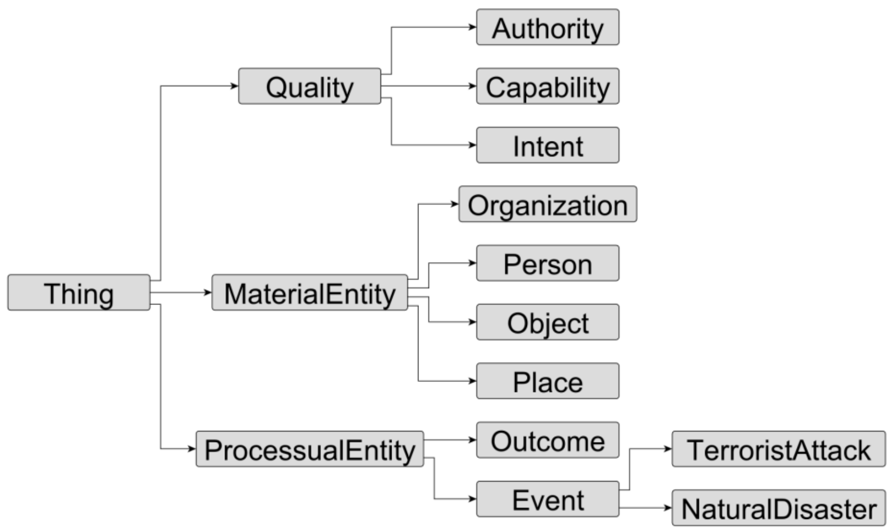

Figure 2. The backbone taxonomy of AIT, with the top three levels and the path to TerroristAttack. Arrows indicate subclass (is-a) relationships.

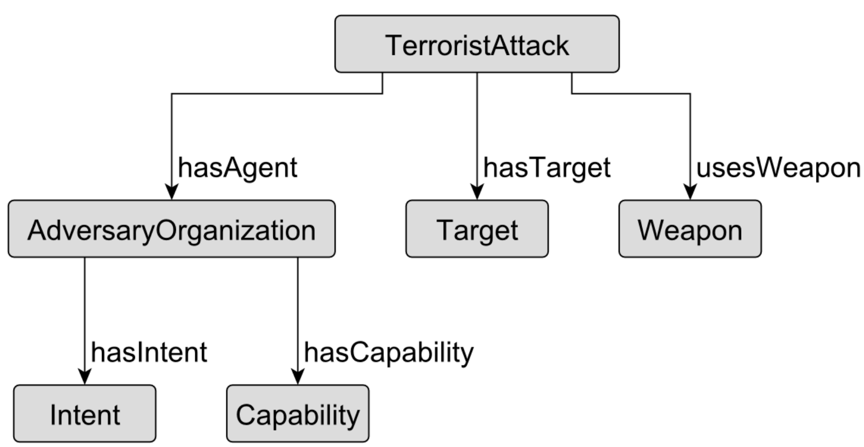

Figure 3. A TerroristAttack requires targets, weapons, and adversaries. Adversaries require capabilities and intents. Arrows indicate various more complex relationships as labeled.

From this backbone we included initial representations of types of Adversaries and their distinctions, and the types of targets and weapons. A particular characteristic of the AIT model was to model the Outcome of a particular attack, and how that might support a particular Intent; this allows some reasoning over which groups might be likely to be implicated in a particular attack, based on their intents and the types of attack. For full details of the model, see Turner, Weinberg, and Turner (2011). A Simple Ontology for the Analysis of Terrorist Attacks. University of New Mexico Electrical and Computer Engineering Department Technical Report EECE-TR-11-0007, http://hdl.handle.net/1928/13714.

4.2 Computational Category Theory

This section describes the theoretical background and algorithms for a software package that implements the basic notions of category theory. In a subsequent effort, the software code will be modified and substantially extended. This effort is an outgrowth of an investigation into a theoretical framework for faceted ontologies equipped with data repositories. There exist other packages developed for category theory and available online; we do not discuss these, except to mention that our impression from an initial investigation of some of them is that they are either not freely available (that is, there could be legal issues involved in using them), are of limited applicability, or are intended mainly for instructional use in mathematics and, again, are of limited applicability. We apologize to those involved with these other packages should it turn out that we have overlooked an available, open- source package that would have met our needs. Another topic that concerns our work is the justification for it. There are many works involving the development of ontologies for computer applications, which after all is the intended outgrowth of our work. Yet another topic is the mathematics itself. Category theory was for many years regarded as ”pure, abstract mathematics”. How, then, do we claim to be applying it? And what is it? We therefore begin with a discussion of ontologies, faceted ontologies, mathematical semantics, and category theory, and how this last item applies to the others.

In anticipation of the discussion to follow, let us begin with some brief facts about category theory. It is the mathematical theory of structure. It deals with a multi-level hierarchy of mathematical systems and, hence, allows the mathematical objects—algebras, geometric spaces, differentiable structures and so forth—to be investigated in relation to each other and at different levels of abstraction. It has seen substantial and increasing use in computer science, where it provides a mathematics for formal semantics, as well as in certain areas of mathematics. Yet, category theory is as yet unfamiliar to many and is often spoken of as a purely abstract, very difficult field of study. There is evidence to the contrary from applications, and we have included what we hope is all the necessary background knowledge to counter the notion that category theory is inaccessible.

The organization of the report is as follows. The remainder of the Introduction provides a background for faceted ontologies and some considerations related to the use of category theory as mentioned in the previous paragraph. Most of this background discussion has been adapted from a previous report. Section 2 provides a basic introduction to category theory and describes the passage from finite graphs to categories. Section 3 introduces further quantities from category theory and describes their use in constructing the full faceted ontology. Section 4 describes the software and discusses the considerations that needed to be addressed in developing it. Section 5 presents the initial test results.

An ontology is an expression of that which exists, that is, objects, properties, events, processes and their relationships in a universe of discourse, which we shall variously call either a world or a domain1. A faceted ontology is a sort of “ontology with faces”, where each face is an ontology specialized to a particular viewpoint on the domain. Any investigation of ontologies and the items they are supposed to express necessarily

involves the symbolic representation of knowledge about things and their relationships in some conception of a world or domain. In the recent application of ontology outside philosophy, it also involves the correspondence of the symbolic representation to data, which we consider in the form of a mathematical model of, say, a computer system such as an online data repository. Coupling an ontology to data gathered from the world it is supposed to describe requires that it have an unambiguous semantics. By semantics we mean the meaning of the terms in a symbolic system, or symbolic structure, as determined by a systematic interpretation of them in the world environment. In the work presented here, this requirement is addressed with mathematical rigor, the idea being to disambiguate the semantics of symbolic representations so that their correspondence with data is accurate and precise. This will facilitate the development of computerized systems that allow the exploitation of ontologies, for example in performing updates to relational data repositories and in gaining useful information from the data.

As suggested in the opening paragraphs, category theory provides a basis for expressing mathematical structure, and this investigation involves the underlying structure of ontologies, systems of ontologies, and data repositories or databases associated with them. This stems from two notions: First, the key to understanding the semantics of a computer system language is to view it as a means of expressing knowledge and applying it through the operations of a computer system. Second, semantics is structure, that is, things and their relationships. In this investigation there are three kinds of structure.

The first kind of structure is that of an ontology. This is expressed as a network of entities, which are either labels for or informative descriptions, or concepts, of different kinds or classes of things, and descriptive links showing how the things in the world or domain that are associated with one concept relate to those of another. This structure can be presented as a graph whose nodes are the various entities, normally represented by suggestive labels, and whose links are relationships between the entities, where the links are also given suggestive labels. The links have a sense of direction from one node to another, for example,

$$dog \xrightarrow{is-a} mammal$$

would be an “is-a” relationship between the class of dogs and the class of mammals indicating that every dog is also a mammal. There is a large body of mathematics devoted to graph theory, of which the graphs discussed here are a special case. The fact is, however, that in computerized ontology work the mathematics is typically absent and the graphs exist purely for visualization. Visualization is a valuable aid to understanding the semantics of a system, and we certainly do not suggest replacing visualization as a tool with a mathematical tool via manipulation of equations or other symbolic formulas. Part of the motivation for the work described here is the notion that graphs are a valuable tool for illustrating the relationships among entities in ontologies, but for the formalization of ontologies they are not as expressive of intended meaning as are categories.

Given that graphs are evidently a straightforward format for ”sketching out” ideas for ontologies, and are after all easily formalizable as mathematical objects, we shall assume that a collection of graphs is given. Because of the graph connectivity and the labeling of

the nodes and links, this collection is regarded as an expression of a faceted ontology. Because categories offer a formalization of ontologies that is more expressive of its semantics, the collection of graphs is to be converted to a system of categories. Explaining how this comes about will entail more mathematical depth than is common in discussing ontologies. Before proceeding further along this line of thought, let us pause to consider why we are pursuing such mathematical depth in the first place. Why invest all this effort when informally drawn graphs and systems such as OWL for developing ontologies on the web are already in place?

Consider an ontology for beverages, part of which is shown in the form of an entity-relationship (ER) graph in Figure 4. The entities Beer, Wine, etc. have is a links to the entity Alcoholic Drinks. The entities Grapes and Grains have is a links to Plants. There is also a made from link from Wine to Grapes and another from Beer to Grains. The terminology is suggestive: The is a type of link has already been defined (informally). The made from link expresses the fact that, for example, wine is made from grapes. It is tempting to say that the meaning of the graph with its labeled nodes and links is clear. However, the meaning behind the node and link labels in the graph is clear only to we humans, and that only because the labels are familiar expressions derived from natural language such as Beer, is a, Grains and made from. Labels alone are insufficient for a computer system for ontology and data manipulation because it must be programmed specifically to enforce the understanding implicit in the labels. For example, in symbolic inferencing with the ontology,

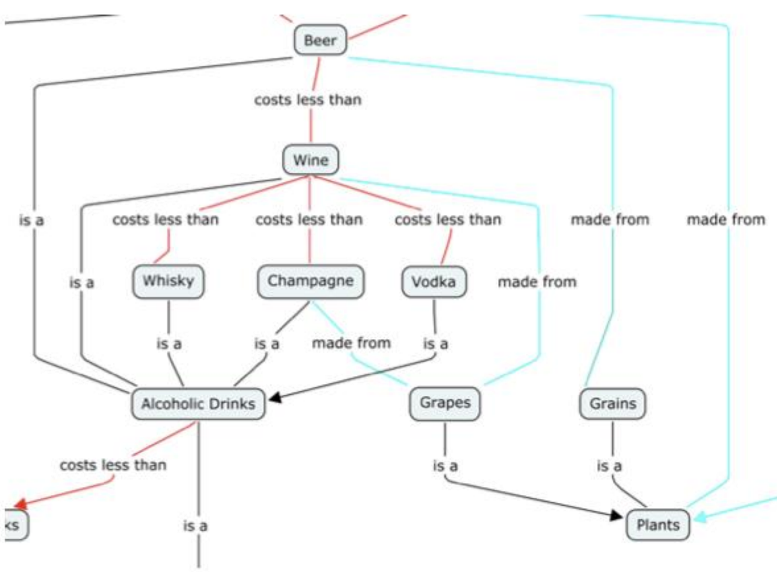

Figure 4. The Beverage Ontology as an entity-relationship (ER) graph.

the computer system can find the entities that are subconcepts of Alcoholic Drinks only if it is programmed for symbolic manipulation and can recognize an “is a” link and associate the symbolic expression with its meaning in terms of the search operation to be performed. The programming involved must convey the semantics implied by the graph, that is, the intended meaning of the structure of labeled entities and links. This requires a translation of symbol system into computer system with a human-friendly interface. The

translation requires some degree of formalization, preferably made explicit in a system specification or at least in a thorough system document once developed. Ontology development systems such as OWL are programmed to recognize properly-formed symbolic expressions and enforce the labeled graph semantics as indicated. The aim of the present effort is to explore the notion that category theory offers a useful mathematical adjunct to such systems and their accompanying visualization tools. With mathematics, the actual semantics implicit in an ontology and any data associated with it, as opposed to the intended semantics, can be clarified. This allows any differences between the actual and intended semantics to be resolved through refinement operations. Further, the correct computer implementation of the intended semantics can be at least partially automated.

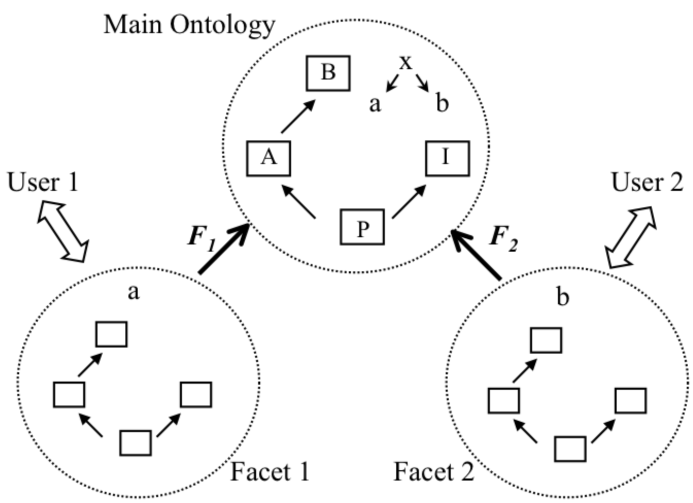

Figure 5. A faceted ontology.

The second kind of structure is the relationship between different but related structures of the first kind. As discussed earlier, the Beverage Ontology concerns different beverages such as wine and beer and also their principle ingredients such as grapes and grains. Suppose that this ontology has users who are specialists of different kinds; one user, such as a brewmaster, might be concerned with how the beverages are made (mixing, fermentation, brewing, etc.), another with marketing and pricing information (sold by the bottle, price varying within such-and-such a range), and another specifying their recommended use (for example, whether for social drinking or to accompany dining, and with which foods if for dining). An all-inclusive ontology is desirable to maintain the organization of all the information present in the graph of Figure 4. The different kinds of users, however, might find the inclusion of all the information associated with each concept at best cumbersome and at worst confusing. Faceted ontologies are meant to address this by providing additional structures that correspond to the concepts and links of the all-inclusive ontology but provide a different interface or facet for each type of specialist. The facets access the information specific to their specialties through the all-inclusive or main ontology. In turn, the main ontology maintains the coherence of the total body of information present; it associates the information for each concept specific to one facet with that for the same or a similar concept in another facet, and associates the

links likewise.

The information in each facet can be expressed in terminology understandable by its type of specialist, while the main ontology maintains the information in a “neutral” language. This is indicated in Figure 5, where the legend “x” with arrows to “a” and “b” inside the circle for Main Ontology indicates that it maintains information for specialties “a” and “b” combined and that this is translated into the specialty languages. Facet 1, on the other hand, is shown as having a node-and-link structure identical to that of the main ontology but containing only information for specialty “a”. If the Main Ontology contains information for how beverages are made, how they are priced and marketed, and their intended use, then Facet 1 contains only the information on how they are made, Facet 2 contains only the information on pricing and marketing, and so forth. The node-and-link structure of facets 1 and 2 may duplicate that of the main ontology, but only facet-specific information is labeled. For example, if Facet 2 is concerned only with pricing and marketing of the beverages themselves and not at all with their ingredients such as grapes, then the concept Wine is shown but the concept Grapes and the is a link from Wine to Grapes need not be labeled as such except in the main ontology. For flexibility, however, it is desirable to include them in some form unobtrusive to the Facet 2 user. For example, at a future time it might become desirable for Facet 2 to include information about the pricing of grapes, since that can contribute to the price of wine. Hence, the ability to have the entire structure but strongly highlight and label only the relevant concepts and links can be useful. Another consideration is the ability to perform data repository updates through the facet interface that are propagatable through its links to other classes.

To summarize, an example of the second kind of structure, a faceted ontology, consists of a system of correspondences between ontologies that are related by subject matter. Figure 2 illustrates the scheme we use for faceted ontologies. It consists of a main ontology joined to the facets, which are ontologies specific to different users’ areas of expertise within the subject matter represented in the main ontology. The arrows labeled F1 and F2 are correspondences between the main ontology and Facets 1 and 2, which are related in that their common domain is the subject of beverages. The main ontology contains the world-view that maintains coherence among the facets; in this scheme, it contains all the information present in both, and perhaps more. The correspondences are mappings of concepts to concepts and links to links and are programmed in the software system that maintains the faceted ontology. The users interacting with their facets can make facet-specific queries that are forwarded to the main ontology via the mappings. The main ontology in turn resolves each query and supplies this information to the facet for access by the user. It does this with the aid of an inference engine—a symbolic manipulation system that traverses the links to infer something associated with one entity based upon items associated with others. Performing the inferencing in the main ontology allows all information to be accessible in deriving answers to queries originating from specific facets.

At this point, let us settle upon our terminology to prevent any confusion that may arise. As the foregoing discussion has perhaps suggested, we shall use the term ”ontologies” to mean either “a single ontology” or “a facet of a faceted ontology” or “a category that

expresses an ontology". The term "faceted ontology" will refer to a system of interconnected ontologies that can be used interoperationally because they represent separate perspectives on a single kind of world or domain. Which usage is meant will be clear from the context.

The third kind of structure is that of a relational data repository associated with an ontology. A repository contains the data for the classes in the ontology and mapping links, or correspondences, which link the data items of one class to those of another as specified by the ontology graph links. As shown in Figure 6, there is a mapping link between data classes in the repository for each symbolic link between concepts in the ontology. A relational database is one example of this kind of structure, where the data items for a concept are represented as a column in a table, and the mapping links as rows. This is a format for expressing a finite set and its correspondences with other finite sets as indicated in the graph. Thus, a mapping link corresponds to a function that maps the elements of one set to elements in the other (this may require some restructuring for an arbitrarily-given graph, but can always be done). Here again the issue of semantics arises if the repository is to ensure the integrity of its data: the concepts and their links must be matched by the appropriate sets of data and mapping links. This correspondence of structures must be maintained for all facets during repository updates performed by a user at a single facet; this corresponds to the view update.

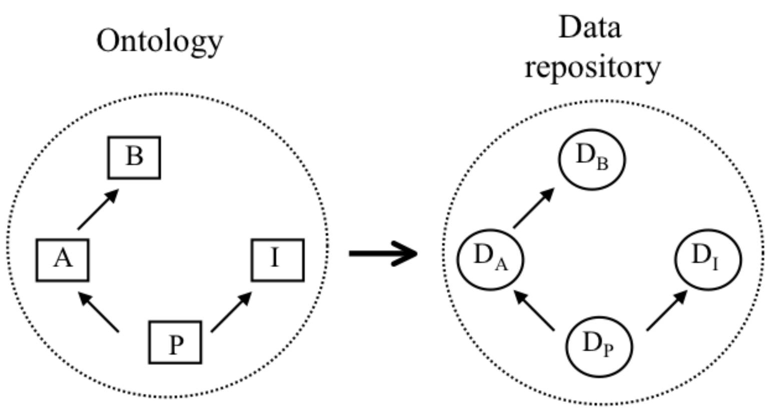

Figure 6. Three structures: an ontology, a data repository, and a correspondence between the two.

The association of the classes and relational links of a repository with an ontology provides an additional example of the second kind of structure. The large arrow between the ontology and repository in Figure 3 has smaller arrows, or maplets (not shown), which associate each concept of the ontology with its associated set of data items (usually presented as a column in a table) and each link of the ontology with a data mapping (a row in a table). As will be seen, these three structures require more information than a graph conveys to properly express the semantics of an ontology and any data associated with it. Equally important is the consideration that updates to the data associated with one facet must be consistent with changes to the data of each other facet. The formalization using category theory not only ensures a correct match of data to ontology, but also provides a mathematically rigorous mechanism for maintaining and updating the data

repository. This is done through structure-preserving associations between different ontology-to-data mappings, where each of the latter represents a data repository state and the associations represent updates, changes from one state to another.

Category theory can be applied to represent each of the three kinds of structure. It supplies the mathematical background for a computer system with which users can perform computations that employ and manipulate the structures. This allows inferencing within and between ontology facets, using not only the "flow" along the relational pathways within the internal structure of each facet, but also the seamless transitions between their concepts and relationships based upon the mappings between the facet structures. It also allows the concepts and relationships in each facet to be associated with the data tables and links in the relational data repositories associated with it. This, combined with the mappings between facets, allows seamless transitions between the data repositories associated with the different facets. Finally, mappings between the facet-to-repository mappings allow repository updates to be performed in a mathematically rigorous fashion that ensures consistency with the mappings between the facets. The computational tool that results from this development thereby exploits the many advantages of a mathematical formalization that is based upon the notion of structure.

4.3 Ontology Tools and Applications

A thorough and detailed review of the ontology representation and algorithms literature was carried out. Several key trends were observed. Ontology representations fall under 3 levels, extensional-level: description of basic objects in the domain and their properties, intentional-level: grouping of objects to form concepts, and meta-level: formal abstraction of concepts and higher order concepts [Guarino et al 2001]. Ontology representations can be expressed in terms of class – relations for semantically structural concepts, actions – processes for semantically temporal concepts, or a combination of both for complex semantic concepts. Interpretation of the representation can be in terms of single models that represent specific concepts or multi models that represents higher order templates to create instances of single models [Guizzardi 2007]. The main algorithms for ontologies are based on the concepts of data abstraction in object-oriented languages, type construction and polymorphism in Lambda calculus, frame-based languages, semantic data models, software formalism, graph models, visual models and temporal models [Wache et al 2001, Qin and Hernandez 2004].

However, ontology engineering has some key challenges. Constructing ontologies requires significant time and resources. Generally, expensive domain experts are hired to break down a domain in to classes, individuals, functions, axioms and rules. We feel that this approach is not feasible for this current project. We also feel that most ontologies are feature-heavy and do not optimally utilize of most of their structure. Hence, while we still want to use the representational power of ontologies to denote semantics, we are seeking lighter-architectures for ontology building.

We have applied ontology-engineering techniques to the problem of organizing information from online conversations. We have introduced the concept of light

ontologies that retain the representational richness of formal ontologies, but are easier to implement and maintain.

Since we are effectively trying to model the entire human knowledge base, we look at other sources in which this knowledge has been incorporated and is easily available. Wikipedia is an obvious example. Wikipedia entries are linked to relevant other entries through hyperlinks, much like regular web-pages. However, Wikipedia hyperlinks tend to be relevant to the local context. A link from page A to page B shows that page B is semantically related to (part of) the content of page A.

Hence, we build light-ontologies by exploiting the links between entities in Wikipedia. [West & Precup, 2009]. Since each Wikipedia entry has links to other Wikipedia entries through hyper-linking within the document space, we can build a topology-map of these links for the whole Wikipedia repository, linking each entry to its most closely related counterparts. The whole repository contains 14 million entries. Building a link graph for 14 million entries will be a computationally challenging task. However, this is something that can be pre-computed and updated at regular intervals. The topology-map consists of links between entries in Wikipedia. Depending on the density of arcs and the topology-dynamics, we identify the main-concepts and the sub-concepts. Analyzing out-going arcs will enable us to understand relationships between concepts.

We tackled two distinct problems: organizing a corpus of scientific publication from the neuroscience domain in to a semantic representation, and analysis of customer service chat transcripts for understanding most common outstanding customer issues. Each problem is described in detail below.

- Our basic aim is to supplement classical ontology architectures with a robust stochastic framework for richer representational power, and easier incorporation of online incremental information retrieval from knowledge sources. Used a restricted subset of several hundred papers already manually curated within the BrainMap database that focus on a specific cognitive construct (attention, to begin with) using a variety of cognitive paradigms. This corpus provides a reference comparison for testing, evaluation, and validation. We mined the corpus to identify key concepts, and to learn relationships between the concepts. The structure of the individual papers was mined using computational linguistic techniques like Latent Semantic Analysis, and papers describing similar themes will be grouped using machine-learning techniques, such as unsupervised clustering.

- The inherent characteristics of online chatting means that businesses could easily connect with their customers help resolve their queries and issues, and disseminate information about the business and its products. Many businesses and organizations have taken it a step further and have implement virtual agents to replace the human customer service agent. The virtual agents have the capability to interpret the most commonly asked questions from the customer, give relevant answers to the queries, or guide the customer to other resources that could handle the customer’s issue. Modeling online conversation can help recognize underlying patterns that have a lot of value for many stakeholders. Businesses can mine the

chat transcripts to understand what are the main issues customers generally face, what are the leading causes of customer dissatisfaction, and what information are customers frequently missing out on. Thus, mining the conversation logs can also provide a wealth of information about new growth areas and opportunities for the business. Hence there is a current need for automated tools for modeling online chat conversations. The specific question we tried to answer was: given a series of chat session between a customer and a virtual agent, create a concept map of the most important issues raised by the customer, the most relevant responses to these issues, and other information most relevant to these issues.

We applied machine learning techniques to organize information contained in neuroscience journal and conference papers. We had a corpus of 350 journal and conference papers in neuroscience subfields like attention and memory. We applied unsupervised clustering on the free text and through a process of pruning, able to identify key concepts contained in the literature. Currently, we are in the process of leveraging CogPo, a neuroscience ontology to automatically annotate the papers in the corpus. We have created a dictionary vector of neuroscience terms from the corpus, and have applied the k-nearest neighbor algorithm, to identify conceptual distance between the papers. These distance metrics then drive the annotation.

We have developed an architecture to enrich chatter bots through learning, representation, and conversation control (Fig.7). This prototype architecture can help a chatter bot engage humans in realistic conversations in customer service situations. We have designed a probabilistic FSA based algorithm to model a conversation as a process that flows through different states.

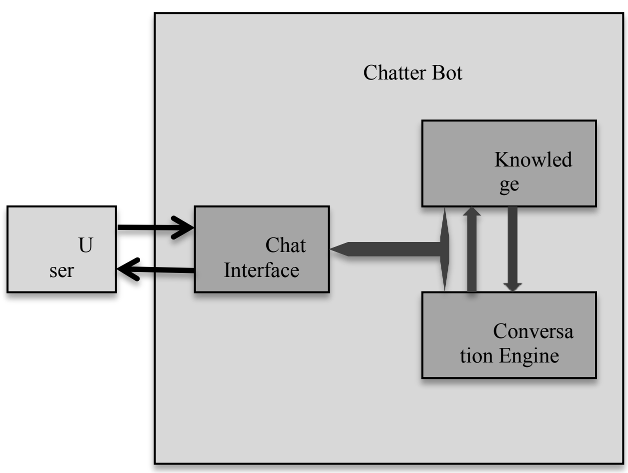

Figure: 7.A chatter bot architecture consisting of a Chat Interface for pre-processing tasks, a Knowledge Engine for representing the domain knowledge, and the Conversation Engine for engineering the direction and flow of the conversation.

4.4 Visualization of Faceted Ontologies

There has been a great deal of work in general graph analysis and visualization over the past decade as there has been lots of work done in ontology development. The challenge in the general work proposed by this project is a combination of quantitative and qualitative scaling, and of aligning the category theoretic features with the represented graph elements.

There are several issues that remain which are specific to *faceted* ontology visualization beyond what has been effected with generalized graph visualization and specifically ontology visualization techniques already developed:

1. Normalization: The development of a generalized facet strategy that can be utilized to describe knowledge structures from any ontology and elaborated from any context.
2. Orthogonality: Method to determine the relationship between two different facets of a merged ontology. Specific methods to help the user visualize these properties will be required.
3. Merging: Describe several facets with some degree of orthogonality by a larger graph entity obtained as the merger of individual facets. Visualization methods to view the whole as well as the parts will be required.
4. Resolution Analysis: The ability to view the data, from merged ontologies, at different levels of specificity. Hierarchical graph visualization methods need to be elaborated further to meet this challenge.
5. Morphing: Starting with one specific facet, a decision maker can "walk" from one faceted context to another by morphing the view and seeing the steps in between. Visualization and user interface methods to move between semantic points of view will be required.
6. Projection: This analysis provides a method to create a new facet by projecting merged graph (result from merging existing facets) into a desired context. Visualization methods for selecting and viewing these facets will be required.

The following describes the requirements and ideas developed while working manually and ad-hoc with hand generated component ontologies, blending ontologies, and the category theoretic diagrams derived from them to create a combined faceted (or blended) ontology.

The unique challenges of this problem include providing one or more representations of individual and blended ontologies that support a range of functions from simple generative editing of ontologies to troubleshooting, to attempting to understand the whole of a blended ontology to drilling down into parts of the ontology from different semantic perspectives.

Of the many tools and techniques developed for complex graph visualization, the two basic approaches we felt were most useful for managing the range of issues and the details of dealing with the building, troubleshooting and utilization of faceted ontologies were hierarchical connection matrix methods and node clustering and edge bundling techniques. Fundamentally, these faceted ontologies require multi-scale management techniques, both to handle overall scale problems and to support detail-in-context.

Currently the team generates ad hoc ontologies first using CmapTools to draw out and discuss a notional structure. CMapTools is a widely used tool for concept mapping and has some useful features. The team then regenerates the same structure (manually) using Protégé which produces an OWL file. This OWL file is then processed to create an appropriate structure for manipulation with Category theory to produce (ultimately) a blended ontology.

The following example is of a hypothetical terrorist attack. Figure 8 is a CMapTools diagram of the expanded detail of a blended, faceted ontology.

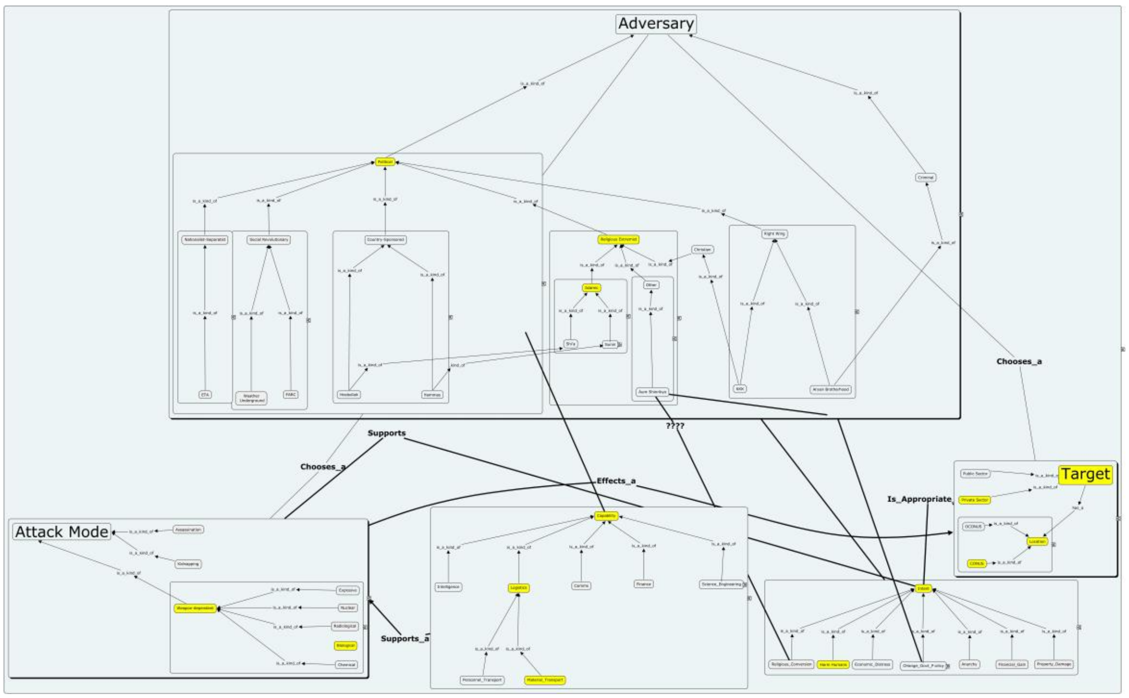

Figure 8. Fully elaborated view of blended facets built with CMapTools. This view shows three levels of detail expanded into a single level.

In this example, the underlying structure of Adversary, Attack Mode, Intent, Capability, and Targets are exposed, including yet more substructure underneath these. For example, one component of Intent might be to change domestic policy… and it may be postulated that certain types of attacks (e.g. radiation) on public (rather than private) targets in the continental united states (CONUS) align with this intent. Similarly, a particular Adversary type with extremist religious views may be proscribed from certain types of attacks (e.g. biological) or targets (civilian populations). All of this is encoded in the links between nodes or subnodes in these ontologies.

Faceted ontologies, as we describe them here, are a combination of ontologies encoded using formalisms from mathematical category theory, with various data sources that are mapped into this framework. These ontologies, formalized as categories can be represented as directed multi graphs. Unfortunately there is no specific precedent for this type of visualization. Drawing on methods for ultra scale, hyper-graphs and multi

graphs, we found Edge Bundling, Node Clustering, and Hierarchical Matrix methods to be the most promising.

We propose to test the sample problems the analysis team is working on by producing prototype tools with the features of the hierarchical matrix and the edge bundling examples above. In both cases, the tools must allow interactive exploration of subsets of the graphs, magnifying the regions of interest without completely obscuring the context of the entire graph.

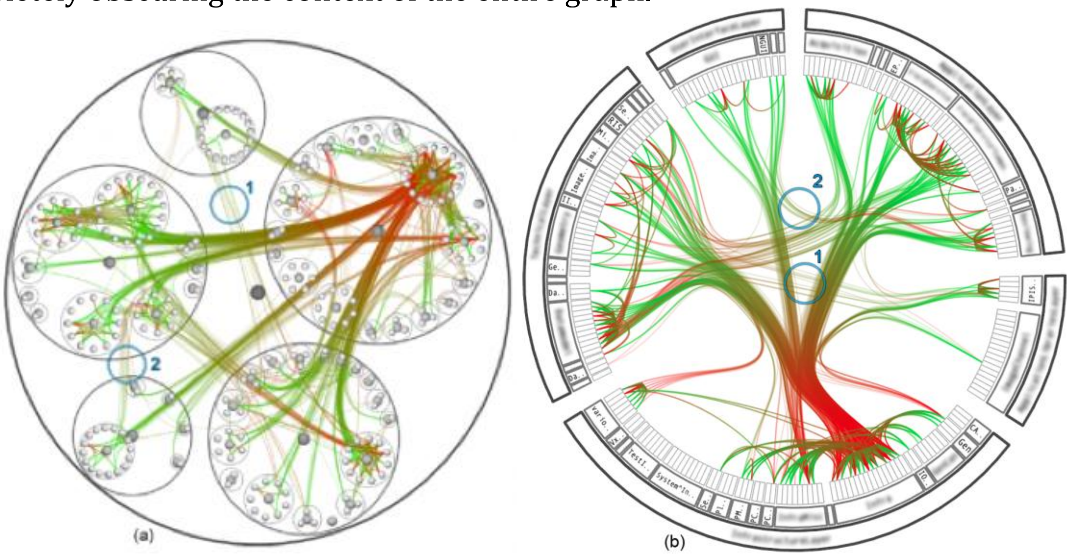

Figure 9. Edge Bundled Bubble (a) and Radial (b) Tree (Holten, 2006)

The first concept in the Edge Bundled Bubble Tree as shown in Figure 9a would be to encode the *is-a* hierarchies in the bubble structure and then the more unstructured relations between elements at different levels with edges which would, when they grow complex automatically bundle using Holten's (Holten, 2006) or similar technique. The user would indicate interest in a subtree, causing it to grow disproportionately to the rest of the tree.

In Figure 9b, an edge bundled Radial Tree, the hierarchy would naturally encode in the radial trees on the perimeter with again edges that would be bundled automatically to expose relations. Specific interest could also be interactively modified, causing segments of the circumference to expand at the expense of others. This focus in context can be applied at different levels of hierarchy as well.

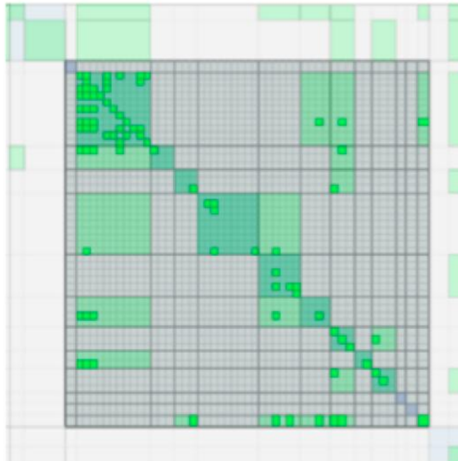

Figure 10. Multilevel or Hierarchical Connection Matrix (van Ham)

The Multilevel or Hierarchical Matrix method would encode hierarchy by grouping elements from each level in the hierarchy together on the axes, yielding a semi-diagonalized matrix with each grouping showing as a denser block along the diagonal of the matrix (Fig. 10). Interconnections between these blocks would show as outliers. By various energy minimization methods, the nodes could be automatically ordered to increase the density along the diagonal. Semantic Zooming could be affected through selection of the regions of interest, causing them to expand at the expense of less interesting (in the moment) regions. Methods similar to those found in Van Ham (Ham, 2003) would be used to hide detail below an appropriate resolution (sub-pixel).

In all cases, we wish to enhance and expose the *facets* in the graph. To begin with, the facets of interest are given a priori by the user who sets up the problem. As a problem grows and analysts study the problem, new facets will emerge or be forced by the user. Not only will we need to be able to manipulate these graph views to aid in that, but we also need to link these methods with the underlying algorithms for blending and morphing between facets.

5. Summary of Significant Accomplishments

This project has six main accomplishments. First, we have successfully formulated the problem of facet ontology merger into faceted ontologies, along with their associated data repository, within the framework of category theory. Second, we have created computer codes that implement the core categorical operations necessary to form categorical faceted ontologies. Third, we have created a series of test cases consisting of simple facet ontologies with domains ranging from beverages to terrorist organizations. These were used to study the consequences of categorical operations on human generated ontologies and to give some level of face validity to the computational results. Fourth, we have successfully demonstrated the application of ontology technologies to automated annotation of documents, and semantic classification of text chat segments. Finally, the fifth accomplishment was to develop a very clear understanding of the range of visual representations for categories as multi-graphs.

These accomplishments are fully documented in the following UNM technical reports (* denotes a student), available on request:

1) "Pre Incident Indicator Analysis (PIIA) System", Frank Gilfeather (UNM Dept. of Mathematics), Thomas P. Caudell (UNM Dept. of ECE), Mahmoud Reda Taha (UNM Dept of Civil Eng) & Dave Weinberg (Practical Risk, LLC.). UNM Technical Report EECE-TR-1-0008, August 17, 2011.

2) "A Simple Ontology for the Analysis of Terrorist Attacks", Matthew D. Turner (Conjectural Systems & NM Mind Research Network), David M. Weinberg (Practical Risk, LLC.), & Jessica A. Turner (UM Dept. of ECE & NM Mind Research Network). UNM Technical Report EECE-TR-11-0007, August 2011.

3) "A Categorical Model for Faceted Ontologies with Data Repositories", Michael J. Healy, Renzo C. Sanchez-Silva* & Thomas P. Caudell (UNM Dept. of ECE). UNM Technical Report: EECE-TR-11-0002, March 21, 2011.

4) "Visualization Techniques for Complex Graph Structured Data", Nate Gauntt* (Dept. of CS) & Steve Smith (Los Alamos Visual Analytics), January 2011.

5) "The initial category theory code for faceted ontologies", Michael J. Healy & Thomas P. Caudell (Dept. of ECE) & Nate Gauntt* (Dept. of CS) (in preparation, May 2012)

6.0 Conclusion

Situational awareness requires acquisition of meaningful and reliable information. In any number of operating environments, large streams of raw information must be analyzed and processed by agencies that range from law enforcement to emergency services during a crisis. This research focused on information related to strategic intelligence collection and analysis. Reports obtained by such processes reveal only pieces of the situational picture – it is the combination of many reports (from different analysts and sources) that potentially reveal the underlying picture. Decision makers will benefit greatly from methods that organize information into new semantic perspectives different from that in which it was collected. This research investigated the organization of context specific information into semantic graphs and the merging of the semantic graphs into a multigraph to create a faceted ontology. This organizes the viewpoint-specific semantic graph structures into a more readily interpretable, robust, perspective neutral representation. The simpler semantic structures are collected from various sources focusing on, for example, socio-cultural networks, geo-spatial distributions, or threat scenario trees. When synthesized into a logical whole, the resulting multigraph or faceted ontology produces a common intelligence picture that gives decision makers insight into the situational roles, goals, relationships, and rules of behavior of relevant groups or individuals. We believe that faceted ontologies will ultimately aid in the discovery of missing informational clues and possibly obscure clandestine activities. This research has created new knowledge in the area of ontology merging, faceted ontologies, inference within them, maintenance of their associated databases, and the visualization of their complex inter-relationships.

7.0 References

Adamek, J., Herrlich, H., & Strecker, G. (1990). Abstract and concrete categories. Cambridge University Press.

Baianu, I. C. (1987). Computer models and automata theory in biology and medicine. In M. Witten (Ed.), Mathematical Models in Medicine–Diseases and Epidemics (Vol. 13, pp. 1513–1578). Pergamon Press.

Barabasi, A-L., 2003, Linked, Plume, New York, New York, 294pp.

Burstall, R., & Goguen, J. (1980). The semantics of Clear, a specification language. In D. Bjorner (Ed.), Proceedings, 1979 Copenhagen Winter School on Abstract Software Specification (pp. 292–332). Springer. (Lecture Notes in Computer Science, Volume 86)

Clark, R. M., 2007, Intelligence Analysis: A Target-Centric Approach, CQ Press, 2nd ed., Washington, DC, 321pp.

Colomb, R., Dampney, C. N. G., & Johnson, M. (2001). The use of category-theoretic fibration as an abstraction mechanism in information systems. Acta Informatica, 38 (1), 1–44.

Crole, R. L. (1993). Categories for types. Cambridge University Press.

Dampney, C. N. G., Johnson, M., & Rosebrugh, R. (2001). View updates in a semantic data modelling paradigm. In Proceedings of the Twelfth Australasian Database Conference ADC2001 (pp. 29–36). IEEE Press.

Ehresmann, A. C., & Vanbremeersch, J.-P. (1997). Information processing and symmetry breaking in memory evolutive systems. BioSystems, 43, 25–40.

S. Eisenhawer, T. Bott, J. Foggia, S. Brown, K. Neitzke and M. Sorokach (2009) Risk-Benefit Analysis of Advanced Air Transportation System Technologies Using Logic Gate Models, Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM2009) Napa, California.

Gangemi, A. (2005). Ontology design patterns for semantic web content. International Semantic Web Conference.

Goguen, J. (1973). Categorical foundations for general systems theory. In F. Pichler & R. Trappl (Eds.), Advances in cybernetics and systems research (pp. 121–130). Transcripta Books.

Goguen, J. A., & Burstall, R. M. (1984). Some fundamental algebraic tools for the semantics of computation–part 1: Comma categories, colimits, signatures and theories. Theoretical Computer Science, 31 (1,2), 175–209.

Goguen, J. A., & Burstall, R. M. (1992). Institutions: Abstract model theory for specification and programming. Journal of the Association for Computing Machinery, 39 (1), 95–146.

Grabo, C. M., 2004, Anticipating Surprise – Analysis for Strategic Warning, University Press of America, Lanham, Maryland, 174pp.

Guarino, Smith and Welty (2001). Ontology: its transformation from philosophy to information systems. Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001

Guizzardi (2007). On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models. Proceeding of the 2007 conference on Databases and Information Systems IV

Gust, H., & Kuhnberger, K.-U. (2005). Learning symbolic inferences with neural networks. In Proceedings of the XXVII Annual Conference of the Cognitive Science Society (CogSci2005). Cognitive Science Society.

van Ham, F., “Using multilevel call matrices in large software projects,” in Proceedings of the IEEE Symposium on Information Visualization, (Seattle, Washington), pp. 29–34, IEEE, Oct. 2003.

Healy, M. J., & Caudell, T. P. (2006a). Ontologies and worlds in category theory: Implications for neural systems. Axiomathes, 16 (1-2), 165–214.

Holten, D., “Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data,” IEEE Trans. Vis. Comput. Graph, vol. 12, no. 5, pp. 741–748, 2006.

Johnson, M., & Rosebrugh, R. (2001). View updatability based on the models of a formal specification. In Proceedings of the 10th International Symposium of Formal Methods Europe (FME 2001, Berlin, March 12-16): Formal Methods for Increasing Software Productivity (pp. 534–549).

Jullig, R., & Srinivas, Y. V. (1993). Diagrams for software synthesis. In Proceedings of KBSE ‘93: The Eighth Knowledge-Based Software Engineering Conference (pp. 10–19). IEEE Computer Society Press.

Khalsa, S., 2004, Forecasting Terrorism – Indicators and Proven Analytic Techniques, Scarecrow Press, Lanham, Maryland, 102pp.

Kolodner, J.L., 1991. Improving human decision making through case based decision aiding. AI Magazine, 12(2):52-68

Kolodner, J.L. 1992. Case Based Reasoning, San Francisco: Morgan Kaufmann.

Lawvere, F. W. (1963). Functorial semantics of algebraic theories. Proceedings of the National Academy of Sciences, USA, 50, 869–873.

Lawvere, F. W., & Schanuel, S. H. (1995). Conceptual mathematics: A first introduction to categories. Cambridge University Press.

Luger, G. F., Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 2009. G.F. New York: Addison Wesley, Section 8.3.3; Kolodner, J.L., 1991.

Luger, G. F., Improving human decision making through case based decision aiding. AI Magazine, 12(2):52-68; Kolodner, J.L. 1992.

Mac Lane, S. (1971). Categories for the working mathematician. Springer-Verlag.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? Journal of Memory and Language, 31, 195–217.

McCue, C., 2007, Data Mining and Predictive Analysis – Intelligence Gathering and Criminal Analysis, Elsevier, Amsterdam, 332pp.

Medin, D. L. (1989). Concepts and conceptual structure. American Psychologist, 44 (12), 1461–1481.

Meseguer, J. (1989). General logics. In H.-D. E. et al (Ed.), Logic Colloquium '87 (pp. 275–329). Science Publishers B. V. (North-Holland).

Mika, P., Oberle, D., Gangemi, A. and Sabou, M. 2004, Foundations for service ontologies: Aligning OWL-S to DULCE. Proceedings of the thirteenth annual WWW conference, New York:ACM Press; Open source JAVA api SOFA: Simple Ontology Framework, http://sopha.projects.semwebcentral.org/

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15 (4), 572–586.

Pierce, B. C. (1991). Basic category theory for computer scientists. MIT Press.

Qin and Hernandez (2004). Ontological representation of learning objects: building interoperable vocabulary and structures. Proceedings of the 13th international World Wide Web conference.

Rosen, R. (1958). The representation of biological systems from the standpoint of the theory of categories. Bulletin of Mathematical Biophysics, 20, 317–341.

Schorlemmer, M., & Kalfoglou, Y. (2005). Progressive ontology alignment for meaning coordination:

An information-theoretic foundation. In F. Dignum (Ed.), Proceedings of the fourth international joint conference on autonomous agents and multiagent systems (pp. 737–744). ACM Press.

Turner, Weinberg, and Turner (2011). A Simple Ontology for the Analysis of Terrorist Attacks. University of New Mexico Electrical and Computer Engineering Department Technical Report EECE-TR-11-0007, http://hdl.handle.net/1928/13714Tversky, A. (1977). Features of similarity. Psychological Review, 84 (4), 327–352.

Tzitzikas, Y. (2002). Collaborative ontology-based information indexing and retrieval (Doctoral Dissertation). Heraklion: University of Crete, Department of Computer Science. Retrieved September 4, 2006, from http://www.ics.forth.gr/isl/publications/paperlink/phd-tzitzik.pdf

Uschold, M., Healy, M., Williamson, K., Clark, P., & Woods, S. (1998). Ontology reuse and application. In International Conference on Formal Ontology in Information Systems (FOIS '98), Trento, Italy, Proceedings.

Vickers, S. (1992). Geometric theories and databases. In M. P. Fourman, P. T. Johnstone, & A. M. Pitts (Eds.), Applications of Categories in Computer Science: Proceedings of the LMS Symposium, Durham 1991 (pp. 288–314). Cambridge University Press.

Wache, Vogele, Visser, Stuckenschmidt, Schuster, Neumann and Hubner (2001). Ontology-Based Integration of Information - A Survey of Existing Approaches. Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing.

West, R., Precup, D. J. P. (2009). Completing wkipedia's hyperlink structure through dimensionality reduction. International Joint Conference on Artificial Intelligence.

Williamson, K. E., & Healy, M. J. (2000). Deriving engineering software from requirements. Journal of Intelligent Manufacturing, 11 (1), 3–28.

Zimmermann, A., Kr¨otzsch, M., Euzenat, J., & Hitzler, P. (2006). Formalizing ontology alignment and its operations with category theory. In B. Bennet & C. Fellbaum (Eds.), Fourth International Conference on Formal Ontology in Information Systems (p. 277- 288). IOS Press.

DEPARTMENT OF DEFENSE

DEFENSE THREAT REDUCTION
AGENCY
8725 JOHN J. KINGMAN ROAD
STOP 6201
FORT BELVOIR ,VA 22060
ATTN: J. REED

DEFENSE THREAT REDUCTION
AGENCY
8725 JOHN J. KINGMAN ROAD
STOP 6201
FORT BELVOIR ,VA 22060
ATTN: R. KEHLET

DEFENSE TECHNICAL
INFORMATION CENTER
8725 JOHN J. KINGMAN ROAD,
SUITE 0944
FT. BELVOIR, VA 22060-6201
ATTN: DTIC/OCA

DEPARTMENT OF DEFENSE
CONTRACTORS

EXELIS, INC.
1680 TEXAS STREET, SE
KIRTLAND AFB, NM 87117-5669
ATTN: DTRIAC

OFFICE OF THE VICE PRESIDENT
FOR RESEARCH
UNIVERSITY STRATEGIC
PARTNERSHIP
MSC02 1660
1 UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NM 87131